

【 資料 11 】

厚生労働科学研究費補助金（食品の安全確保推進研究事業）
分担研究報告書

ネット上における風評とリスクコミュニケーションの分析

研究分担者 乾 健太郎
（東北大学大学院情報科学研究科 教授）

研究分担者 岡崎 直観
（東北大学大学院情報科学研究科 准教授）

研究要旨

福島第一原子力発電所の事故以降、一部の消費者が福島県及び近隣地域の食品を避けるという動きが見られる。本研究では、福島県産の桃に関するツイート約 29 万件（収集期間：2011 年 3 月～2013 年 8 月）を自然言語処理およびネットワーク分析の 2 つのアプローチで解析した結果を報告する。福島県の桃の購買に関して肯定的な消費者、否定的な消費者のクラスタを発見し、これらのクラスタの成長過程、クラスタ内外での議論の推移、否定派の主張の分析を通じ、福島県産の桃に関する風評の実態とその対策について考える。さらに、震災発生直後の厚生労働省の公式ツイッターアカウント（@MHLWitter）の発言内容を振り返り、災害時に於ける SNS を用いたリスクコミュニケーションの在り方を検討する。

A. 研究目的

東日本大震災に関連して発生した東京電力福島第一原子力発電所の事故では、放射性物質が環境中に放出され、被災地は放射能汚染にも苦しむことになった。事故直後から福島県を中心とした東北地方、関東地方の農林水産物を買控える動きが消費者の間で広まり、農業・漁業関係者は今も深刻な打撃を受けている。

アンケートやインタビューを通じて、福島県産の農産物に対する消費者の意識や行動の現状を調査し、農業政策への提言を行う研究が発表されている。一方で、誤情報や風評の拡散の温床とされたソーシャルネットワークで、福島県産の農林水産物に対する消費者の日常的な意識を分析した研究は、我々が知る限り存在しない。

本研究では、福島県の特産品である桃に焦点を当て、東日本大震災直後から 2013 年 8 月までの Twitter 上での投稿を自然言語処理およびネットワーク分析の 2 つのアプローチで解析する。福島県の桃の購買に関して肯定的な消費者、否定的な消費者のクラスタを発見し、これらのクラスタの成長過程、クラスタ内外での議論の推移、否定派の主張の分析を通じ、福島県産の桃に関する風評の実態とその対策について考える。

現状の風評の実態をデータから踏まえたうえで、震災発生直後の厚生労働省の公式ツイッターアカウント（@MHLWitter）の発言ログを振り返り、災害時に於ける SNS を用いたリスクコミュニケーションの在り方を検討する。

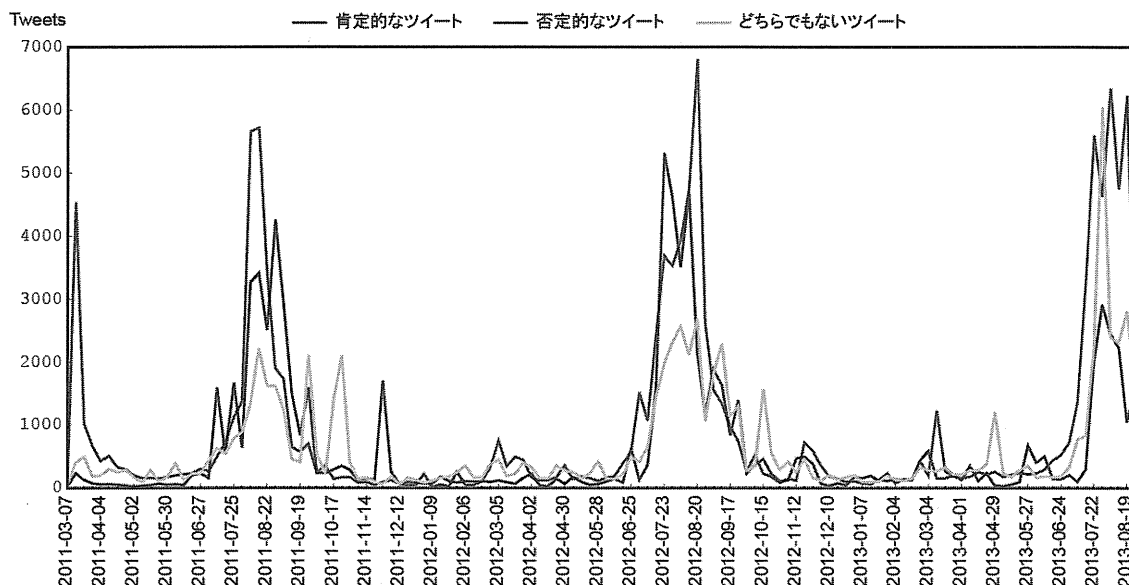


図 1 肯定的・否定的・どちらでもないツイートの数の週ごとの推移。

B. 研究方法

本研究では、2011 年 3 月 11 日～2013 年 8 月 31 日の期間で「福島」かつ「桃 or もも or モモ」を本文中に含むすべてのツイート（425,819 ツイート）を収集した。「ももクロ、福島ハワイアンズでライブ」など、福島の桃に言及していないツイートや、ショッピングサイトへの誘導ツイート、ボット（自動投稿システム）によるツイート等をノイズとして除去し、残った 289,525 ツイートを分析した。

本研究では機械学習に基づく極性分析を行い、全てのツイートを福島の桃を買うことに「肯定的」なツイート、「否定的」なツイート、「どちらでもない」ツイートの 3 カテゴリに分類し、「肯定的」「否定的」の双方の意見を集約した。機械学習器として多クラス・ロジスティック回帰を用い、単語ユニグラム（ツイート中の単語）、単語バイグラム（ツイート中の単語の接続）、風評表現辞書とのマッチ、ツイートを発信したユーザ名、プロフィールの単語ユニグラムから極性を推定した。こうして構築した極性分析器の正解率は 74.27%であった（10 分

割交差検定による）。

また、Twitter 上のユーザの議論の状況を調査するため、リツイート・ネットワークの分析も行った。この分析により、有用なツイートを発信しているユーザや、多くのリツイートをしているユーザ、近い意見を持つグループ（クラス）を発見し、福島の桃に関する代表的な意見の形成プロセスを探ることができる。

リツイート・ネットワークの構築方法は以下の通りである。あるユーザ A のツイートを別のユーザ B がリツイートしたとき、ユーザ A からユーザ B へリンクを作成する。リンクの重み（強さ）はユーザ B がユーザ A のツイートをリツイートした回数と定義する。例えば、ユーザ B がユーザ A のツイートを 2 回リツイートした場合は、ユーザ A からユーザ B へリンクを作り、その重みは 2 である。

このような処理を、今回収集したすべてのツイートに対して適用すると、ユーザをノード（点）、リツイート関係をリンク（矢印）としたネットワークが出来る。このネットワークを Gephi というソフトウェアを用い、ネットワークの可視化を行う。さ

否定的な意見が強い ————— 肯定的な意見が強い



図 2 リツイート・ネットワーク（全期間）

らに、極性分析の結果を利用し、各ユーザー（ノード）の立場を色付きで表現した。

C. 研究結果

図 1 に「肯定的」「否定的」「どちらでもない」ツイート数の週ごとの推移を示した。毎年、桃の出荷時期となる夏に、ツイート数のピークが見られる。2011 年と 2012 年は肯定的と否定的なツイートの数がほぼ拮抗しており、2013 年は肯定的なツイートの方が優勢的であった。2011 年夏の主なツイートは、農家の写真と共に桃の安全性をアピールするもの、桃農家に向けられた冷たい視線に対する悲しみ、福島の桃の当たり年、産地隠蔽の懸念、政治家の不適切発言、市民団体によるセシウム検出であった。2012 年夏の主なツイートは、除染の努力によるセシウムの不検出、輸出した福島産桃の完売のニュース、産地偽装の噂・疑い、風評対策プロジェクトへの批判などであった。2013 年夏のツイートは、「桃の涙」というリキュール、両陛下が福島の桃を召上がった話題、ヨウ素検出という誤情報、風評対策プロジェクトへの批判などであった。

図 2 に可視化したリツイート・ネットワークを示す。ネットワークは左側と右側の 2 つの密なクラスタに分かれ、その間を少数のユーザが繋ぐという構造をしている。しかも、左側のノードは赤色、右側のノードは緑色で彩色されていることから、左側のクラスタが否定的なグループ、右側のクラスタが肯定的なグループに分かれている。このことから、福島の桃に関する議論は肯定派と否定派に二分されており、各グループ内で密な情報交換が行われているが、否定派と肯定派の間で議論が交わされていないことが分かった。

D. 考察

福島の桃に関して否定的なツイートの中で、桃の購入に否定的な理由は以下に集約された。

- （暫定）基準値に対する不信感
- 0 ベクレルへの拘り
- 基準値以下だから安全というのではなく、検出結果の数値を提示してほしい
- 国・東電が補償すべきで、消費者がリスクを負うのはおかしいという主張
- 安易な風評対策に対する批判

否定的な意見が強い ————— 肯定的な意見が強い

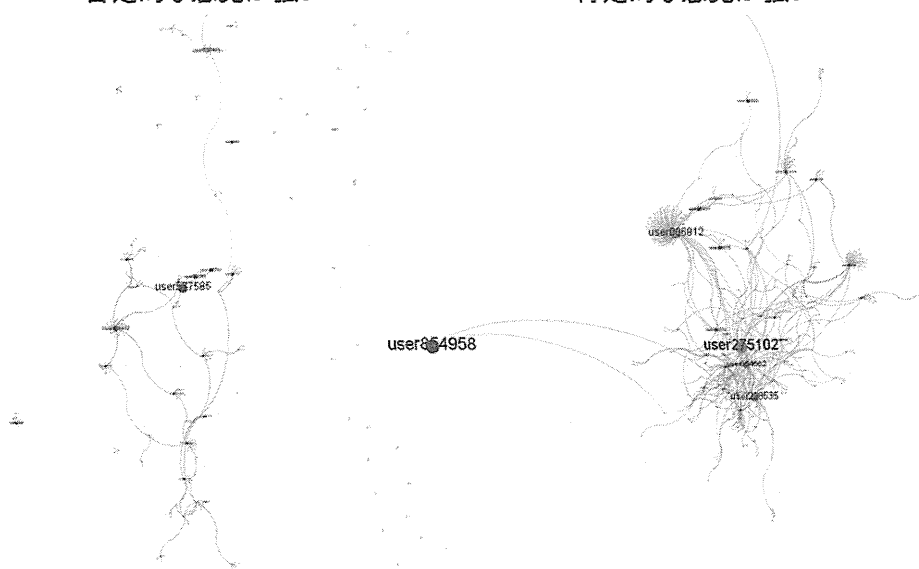


図 3 リツイート・ネットワーク (2013 年夏のみ)

政府が定めた基準値への不信感が根強く、単に「基準値以下」とか「安全」という説明では納得できないという意見が多く寄せられていた。桃の購入に否定的な理由の部分でも述べたように、検査結果をもっと提供して欲しいという意見もあるので、食品のセシウム濃度の基準値の根拠の説明や、検査体制・検査結果の透明性のアピールがさらに必要であると感じた。

図 2 では肯定派と否定派がはっきりと分かれたユーザの空間が可視化されていたが、このネットワークを 2011 年 3 月以降で短い期間（例えば 1 か月）に区切って描画すると、福島に関するクラスタの形成プロセスを動画として確認できる。その動画によると、肯定派・否定派のクラスタは 2011 年夏頃に形成されており、その後、クラスタ内の人々が自分の立ち位置を反対側に変えるような動きは見られなかった。これは、福島の桃に対する人々のスタンスは事故直後に固まってしまい、風評対策は後手に回ると困難を極めることを表す。

ただ、2013 年には特異な動きも見られた。図 3 に、2013 年夏のリツイート関係に絞って可視化したものを示した。まず、2013

年では否定側のクラスタの規模が縮小する傾向が見られた。これは図 1 のツイート数の分析結果に一致する。また、図 2 では否定派の中心的な役割を果たしていた user854958 氏が、図 3 では否定派から切り離されている。これは、user854958 氏が危険情報に反射的に飛びつく人々から距離を置こうとしたためで、氏の立ち位置が変化した訳ではない。一方で、user854958 氏は自身が設定した基準値未満の食品は産地を問わず（福島県産であっても）食べることも述べている。このように、否定派の中でも福島県産の食品に対して多様な考え方があり、検査体制・検査結果に関するデータがもっと開示されれば、風評を払拭できる可能性がある。これは、風評対策を考えるうえで興味深い事例と言えよう。

この知見を踏まえつつ、震災発生直後の厚生労働省の公式ツイッターアカウント (@MHLWitter) の発言内容を以下の観点から振り返ってみた。

1. Twitter ツールの活用意義・意図は？
2. 利用者を理解し情報が伝わりやすいツール・設計になっているか？
3. 利用者のニーズを把握した上での情報

発信か？

4. 発言に対して根拠や詳細説明、具体策の提示を行っているか？

公式ツイッターアカウントのプロフィール欄には「厚生労働省では、ツイッターを通じて国民の皆様向けの情報を発信しています。ツイッターに寄せられたコメントへの返信はしていません。」との記述があり、厚生労働省と国民との双方向のコミュニケーションを想定していないことが伺える。しかしながら、国民からは図4(a)のようなレスポンスが得られており、Twitter というメディアを使って意見を述べようとした形跡がある。この声を拾う体制が整っていないのであれば、Twitter を用いた情報拡散の価値が大きく損なわれてしまう。実際、図4(b)のように、Twitter で広報活動を行うのであれば、PDFによる情報配信を避け、テキストコミュニケーションに最適な配信形式を選ぶべきだという意見が早い段階から寄せられていた。情報拡散にTwitter を活用する意義を十分に検討した上で、メディアの特性を活かした配信方法で情報公開を行うべきであろう。

また、図4(c)のように、国民の情報ニーズと広報内容の間にズレが生じているケースも見受けられた。この場合、国民の関心事（つまり安全面）についての情報発信がされていないことにより、国民からは一方的な発信に見え、不信感を与えている。さらに、図4(d)では具体策が示されていないため、配信した情報に対して多くの批判が寄せられている。これらのケースは、情報のニーズや開示する情報の詳細度により、情報やその発信者の信頼度が変化する事例と捉えることができる。災害という緊急時に的確な情報発信を行うことは難しいが、人々の意見や考え方は早い段階で凝り固まってしまうことを考えれば、日常的に情報を開示し、厚生労働省が配信する情報の信頼性を高めておくことは、風評を軽減する

ことに効果的なのではないかと考える。

E. 結論

本研究では、Twitter 上での福島県産の桃に対する立場と、それぞれの立場における議論を分析した。福島県産の桃を肯定的に捉える立場と、否定的に捉える立場に分断されており、立場を途中で変えたり、立場を横断して健全な議論をした形跡が見られないことから、風評に対する対応の難しさが改めて浮き彫りになった。ただ、否定派の中にも、産地ではなく精密な情報を使って自ら判断したいという意見があり、検査結果に関する数値データを積極的に開示していくことで、両派の溝を解消する可能性を垣間見ることができた。さらに、震災直後に厚生労働省の公式ツイッターアカウントから発信された情報を調査し、SNSを情報拡散手段として用いる場合の課題を報告した。

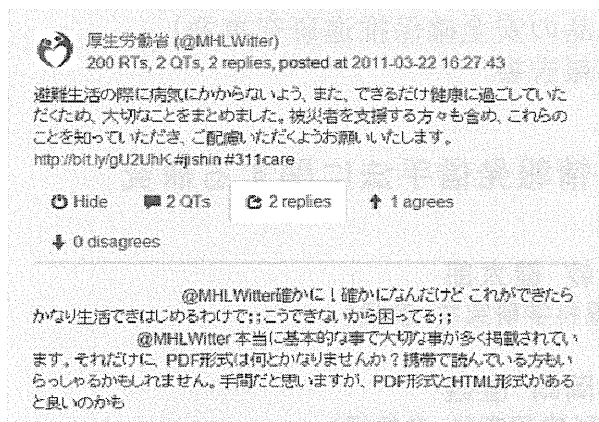
F. 研究発表

1. 論文発表

- 鍋島啓太，渡邊研斗，水野淳太，岡崎直観，乾健太郎．訂正パターンに基づく誤情報の収集と拡散状況の分析．自然言語処理，Vol. 20，No. 3，pp.461-484，2013年6月．

2. 学会発表

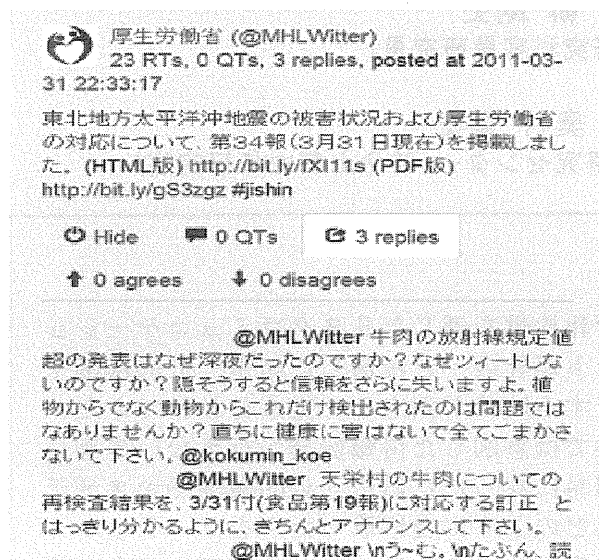
- 岡崎直観，佐々木彬，乾健太郎，阿部博史，石田望．ツイッター分析に基づく福島県産桃に対する風評の実態解明とその対策．第26回日本リスク研究学会年次大会，B-5-3，2013年11月．



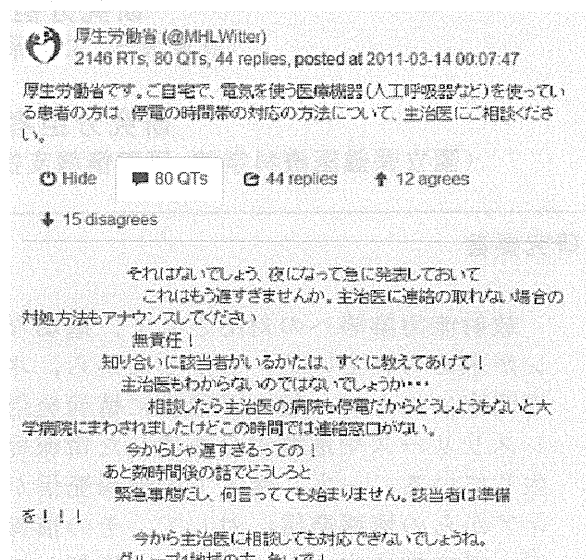
(a) 国民からのフィードバックの例



(b) 情報発信の形式に関する提案



(c) 厚生労働省と国民の情報ニーズのずれ



(d) 具体策や詳細説明の必要性

図 4 厚生労働省の公式ツイッターアカウントからの情報発信と国民の反応

【 資料 12 】

厚生労働科学研究費補助金（食品の安全確保推進研究事業）
分担研究報告書

ソーシャルメディアにおける情報発信手法に関する研究

研究分担者 乾 健太郎
(東北大学大学院情報科学研究科 教授)

研究分担者 岡崎 直観
(東北大学大学院情報科学研究科 准教授)

研究分担者 榊 剛史
(東京大学 工学系研究科客員研究員)

研究分担者 奥村 貴史
(国立保健医療科学院 研究情報支援研究センター 特命上席主任研究官)

研究要旨

放射能汚染等への対応として、適切な情報提供を通じたリスクコミュニケーションが求められている。しかしながら、ネットが社会へと浸透した結果、公的機関や専門家に加えて一般人が独自に情報発信をする時代が到来し、従来行われてきたプレスリリースや地域集会を通じた情報提供では意図した情報伝達が行えない事態が生じている。一方、個人での情報発信が容易になったために、公的機関やマスメディアからの情報発信に対して、その情報をどのように受け取ったかをソーシャルメディアに投稿するという個人が増加している。つまり、ソーシャルメディア上の情報を分析することで、公的機関やマスメディアが発信した情報が不特定多数にどのように受け取られたかを解析できる可能性がある。情報の受け手の反応が明らかにすることができれば、それをフィードバックとして情報提供・発信をより適切な方法に修正していくことが可能となると考えられる。

本研究分担では、ソーシャルメディアのアカウント及びウェブサイト上でのプレスリリースについて、より効果的に情報を発信するための知見を得ることを目指す。具体的には、ソーシャルメディア上の投稿を分析することで、1) ソーシャルメディア上での情報発信についての現状、2) ニュース記事の表現が受け手に与える影響、3) ニュース記事とその反響の大きさの相関について分析を行った。1) においては、ソーシャルメディアのアカウントによる情報発信の効果を測定するために、政府・省庁のソーシャルメディアアカウントの情報発信に対する調査を行った。また、ウェブサイト上のプレスリリースにおける効果的な情報発信手法を検討するために、2) 発信する情報に含まれる感情的表現が受け手に与える影響、3) 発信する情報の言語的特徴が反響の大きさに与える影響について分析を行った。

分析の結果、ニュース記事の感情が受け手の感情に影響をあたえること、ニュース記事に含まれる言語的特徴が反響の大きさに影響することを明らかにした。本研究分担の知見を活用し、ウェブ及びソーシャルメディアでのリスクコミュニケーションにおいて、より効果的な情報発信方法を模索していくことが期待される。

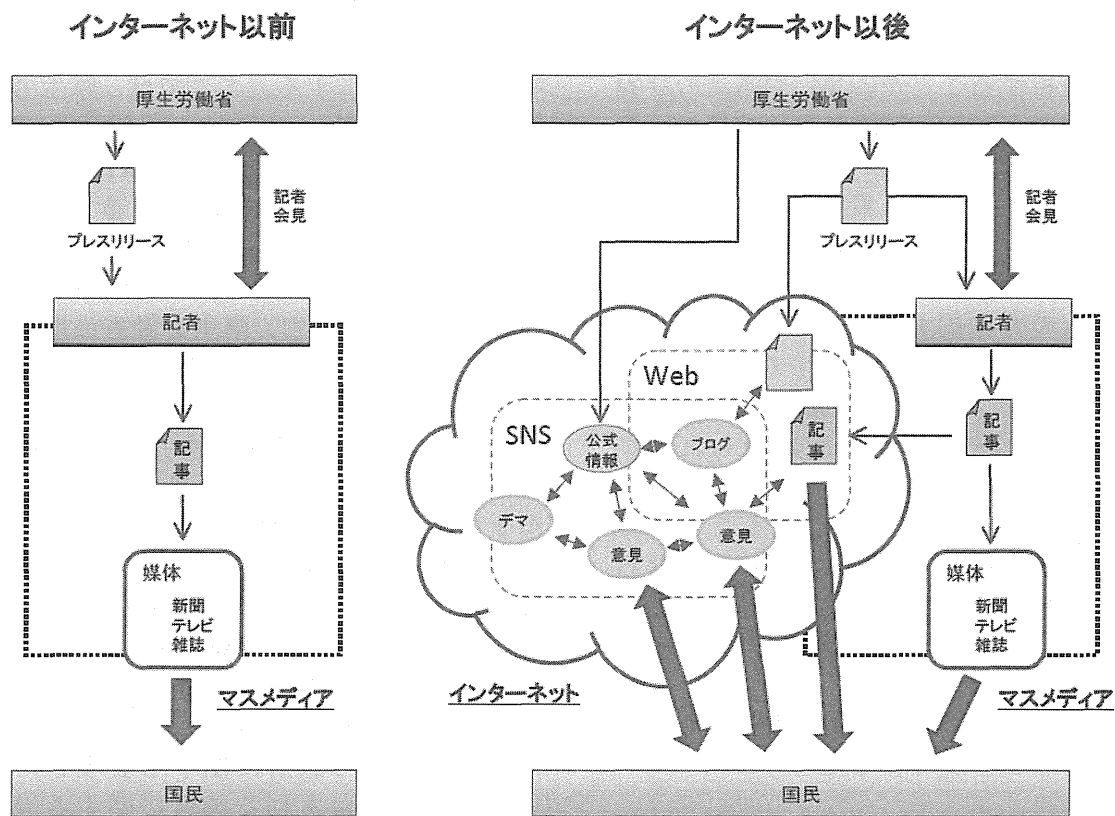


図 1: ソーシャルメディア登場前後における情報発信手法の変化

A. 研究目的

2011 年の東京電力福島第一原発事故後より、放射能に関する正誤のあいまいな情報が数多く生じている。「日本の食品中の放射性物質の基準値が欧米のものよりも遥かに高い」、「福島を中心に健康被害が続出している」、「原発作業員に多発している急性白血病が隠蔽されている」等は、代表的なものであろう。狭い国土において生じた未知の災害に対して社会不安が生じること自体は避けがたいことではあるが、その結果、安全性が確認された農産物等に対しても忌避が進み、多大な社会的損失を引き起こしている。そこで、公的機関に対して、放射能に関する適切な情報提供を通じて無用な社会的損失を避け、被災地の復興を支援していくことが望まれている。

このような「リスクコミュニケーション」

は、今までにも、環境汚染による風評被害などの際に実践されてきた。しかしながら、従来のリスクコミュニケーションは、マスメディアを通じた一方通行な情報伝達か、専門家と一般人の顔の見える交流かのいずれかを前提としており、近年のようにネットを介して公的機関や専門家に加えて一般人までもが独自に情報発信をするような状況を想定していなかった。そのため、今後の公的機関による情報伝達においては、ネット上での反応を考慮せざるをえないが、ネットには膨大な情報が存在するため、社会心理学や統計学のみに基づく従来の社会調査手法では状況の全体像を把握することすら困難となっている。このように、ソーシャルメディアの登場以前以後の情報伝達経路の変化を図 1 に整理する。

一方、個人での情報発信が容易になったために、日常的な行動や意見をソーシャル

メディアに投稿する個人が増加している。そのため、ソーシャルメディア上の個人の投稿を分析することで、CM や商品の評判を測定したり、株価変動や選挙結果の予測を行う試みが増加している。それらソーシャルメディア上の個人投稿の中には、公的機関やマスメディアからの情報発信に対して、その情報をどのように受け取ったか、というものも存在する。つまり、ソーシャルメディア上の情報を分析することで、公的機関やマスメディアが発信した情報が不特定多数にどのように受け取られたかを解析できる可能性がある。こうした情報の受け手の反応を明らかにすることができれば、情報発信の効果測定が可能になると考えられる。もしそのような効果測定が可能になれば、情報発信のあり方について仮説検証のサイクルを回すことができるようになり、より効率的な情報発信の方法を模索することができるだろう。

そこで本研究分担では、ソーシャルメディアアカウントの投稿やウェブ上のニュース記事に対するソーシャルメディア上のユーザの反応を分析し、ウェブサイトやソーシャルメディアでのリスクコミュニケーションにおいて、より効果的な情報発信手法を開発するための知見を得ることを目的とする。

まずはソーシャルメディア上での情報発信について、それに対するユーザの行動調査を行い、ソーシャルメディア上でより効率的に情報発信する手法を模索する。そのために、政府・省庁の Twitter アカウントの投稿やユーザとのやりとりを分析し、各アカウントの露出量や波及効果を測定するとともに、厚生労働省と他省庁のアカウントの比較を行う。さらに厚生労働省が関係する 11 件のトピックについてツイートを収集し、それらのトピックについてソーシャルメディア上でどのような議論が行われているかを明らかにする。

次に、発信された情報の内容が情報の受け手にどのような感情を与えるかを分析

し、ウェブサイト上のプレスリリースにおいて、より好意的に受け取られる情報発信手法を明らかにする。ここでは、厚生労働省のプレスリリースに対する反応を分析対象とする予定であったが、Twitter のデータ取得 API の利用制限上、過去のプレスリリースに対する検索が困難であり、分析を行うために十分なツイート数が取得できなかった。そこで、厚生労働省に関連したニュースでデータが取得できたものについては詳細な分析を行いつつ、同時に、一般のニュース記事を対象とした分析を行った。具体的には、日本最大級のポータルサイト Yahoo! JAPAN 内のニュースサイト “Yahoo!ニュース” を対象に選び、Yahoo!ニュースのトピックスに掲載された記事を対象として、Twitter におけるニュースコンテンツに対するコメント発言(関連ツイート)を収集する。これらのコメントに含まれる感情に関する表現とニュース記事中に含まれる感情に関する表現とを比較し、それらの相関関係を分析する。

最後に、発信された情報の内容が反響の大きさにどのような影響を及ぼしているかを分析し、ウェブサイト上のプレスリリースにおいてより多くの人に興味を持たれるような情報発信手法を明らかにする。この解析に際しては、受け手に与える感情の分析と同様に、“Yahoo!ニュース” を対象とする。あるニュース記事の関連ツイート数の伸びをその記事の反響の大きさと捉え、ニュース記事に含まれる表層的・言語的・時間的・環境的特徴からコメント数の伸びを機械学習により回帰・分類することで、ニュースの内容が反響の大きさをどれだけ予測できるかを示す。

上記 3 つの分析を通じて、ウェブ及びソーシャルメディアでのリスクコミュニケーションに資する、より効果的な情報発信に関する知見の獲得を目指す。

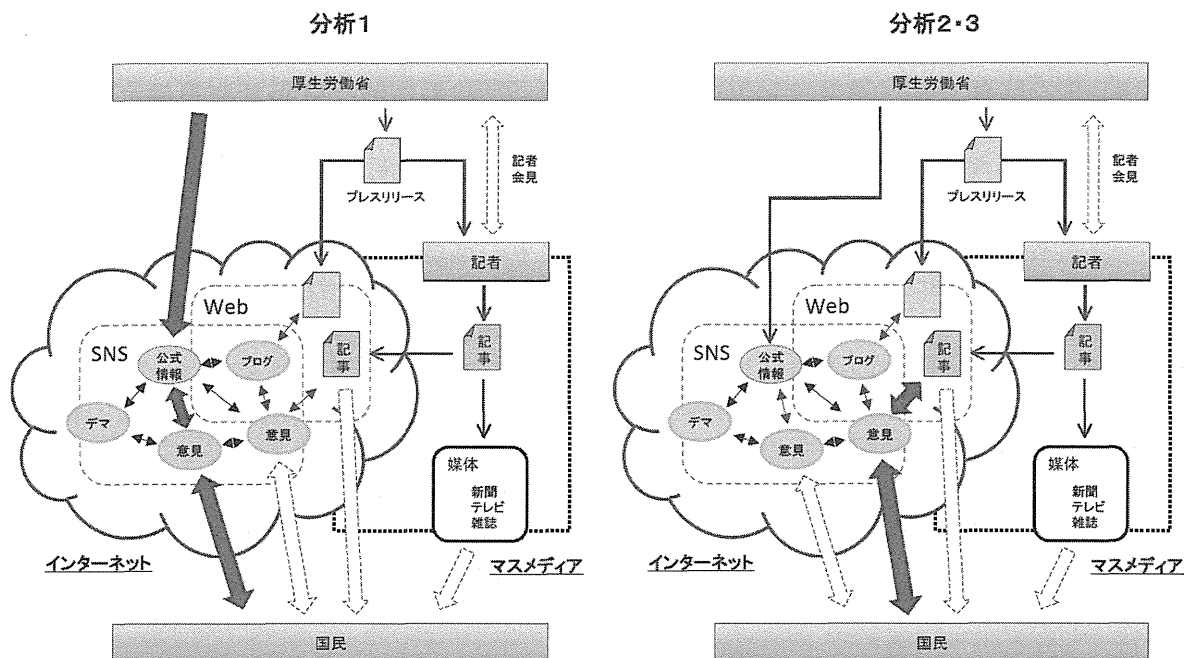


図 2: 本研究が対象とする範囲

これら 3 つの分析の対象範囲を図 1 にならって模式図で表したものを図 2 に示す。1 つ目の分析においては、ソーシャルメディア上で厚生労働省アカウントによる発信情報とそれに対するユーザの反応を分析対象としている。それに対し、2 つ目、3 つ目の分析においては、プレスリリースを用いた情報伝達経路のうち、マスメディアによる発信情報とそれに対するユーザの反応の分析対象としている。つまり、厚生労働省からみて、分析 1 が扱う情報は直接的に変更可能な対象であるが、分析 2、3 については間接的にしか変更できない対象である。研究としては、分析 2、3 においてもプレスリリースとインターネット上の反応との関係性を明らかにすることが望ましいが、データ収集の制約上十分なデータが収集できなかったため、本稿ではデータ収集が可能な範囲に対象を限定した。全体を対象とするためには、今後長期間に渡る継続的なデータ収集が必要となるであろう。

本稿の構成は以下のとおりである。B 節にて、本研究分担で用いた研究方法について

て説明する。C 節にて、それらを用いてデータ分析を行った結果を提示する。D 節にて、得られた分析結果について考察する。E 節にて本研究分担の結論を総括する。

B. 研究方法

本研究分担で用いる分析手法について述べる。

B.1 ソーシャルメディア上での情報発信の現状

最近、政府・省庁が Twitter や Facebook などのソーシャル・メディアを活用した広報を進めているが、広報の効果を政府・省庁を横断的に比較対照した研究は知られていない。本研究分担では、厚生労働省を含めた政府・省庁の Twitter アカウントの投稿やユーザとのやりとりを分析し、各アカウントの露出量や波及効果を測定することで、ソーシャルメディアを用いたリスクコミュニケーションの現状分析および有効な活用方法について述べる。

本研究分担では、2013年3月1日から2015年1月10日までの期間で、表6に挙げた41件の政府・省庁アカウントから発信されたツイート、これらのツイートに対するリツイートやメンション(返信)を収集した。これらのアカウントから発信されたツイートの総数は31,339件、これらのツイートに対するリツイートの総数は814,88件(*), メンションの総数は111,434件(*)であった。ツイートの収集は東北大学の研究グループが独自に行なったもので、ツイート収集APIの利用規約の制約により、アスタリスク(*)付きのツイート数は実際よりも少ない可能性がある。

このデータを用いて、政府・省庁アカウントの活用状況や効果の分析を行う。具体的には、これらのアカウントから発信されたツイートの数やフォロワー数などの基礎的な統計量に加え、これらのアカウントのツイートがリツイートされた回数を調査する。この調査にもとづき、各アカウントの情報拡散能力や国民とのコミュニケーション能力を分析する。さらに、イクメンプロジェクト、エボラ出血熱、子宮頸がん予防ワクチンなど、厚生労働省が関係する11件のトピック(プロジェクトや問題)をキーワードに含むツイート6,581,461件も収集した。

この11件のトピックに関するツイートを収集する際に用いたクエリを表7に示す。こちらのデータは、厚生労働省が関係するトピックが誰によってどのように議論されているかを明らかにするためのもので、ツイートの発信源や厚生労働省への言及回数などを調査・分析する。

B.2 ニュース記事が受け手に与える感情の相関

次に、ニュース記事が受け手の感情に与える影響について相関分析を行う。近年ニュースに対する人々の反応に関わる何らかの知見を求める声は多く、社会的にもそれらの研究の必要性は高い。そのような研

究の一つとして、ニュース記事のコメント内に「笑う」「怒る」「悲しい」「苦しい」のような感情的な言葉があるほど、読み手に共感と呼ぶことが研究成果として発表されている[2]。上記の研究結果を踏まえ、本研究分担ではニュース記事の感情が受け手の感情に与える影響を分析する。当初は、分析対象として厚生労働省のプレスリリース及び厚生労働省に関連するトピックのみを用いる予定であったが、Twitterのデータ取得APIの利用制限上、過去のプレスリリースやトピックに対するツイート検索が困難であり、分析を行うために十分なツイート数が取得できなかったため、Yahoo!ニュース上の一般のニュース記事を対象とした。

B.2.1 ニュース記事に対するコメントの収集

本研究分担ではYahoo!ニュースにおいてトピックスとして取り上げられた記事を対象とする。Yahoo!ニュースのトピックスとして扱われる記事はヤフーの編集者によって選択され、新たに記事の内容に合うおよそ13文字の見出しが付与され、Yahoo!ニュースの画面で表示される。以上の事柄を考慮し、記事のデータとしては、ヤフー側で付与された見出し、関連リンクの見出し、関連リンク名、記事配信社の付与した見出しと記事全文の収集を行う。

ソーシャルサイトTwitterにおけるニュース関連ツイートの収集には、Twitter Search APIを使用する。1つのニュース記事あたりトピックス画面と記事全文画面の2ページがあり、どちらのページ経由でツイートすることも可能であるため、対象とするニュース記事のヤフー作成の見出しまたはトピックス画面へのリンクが含まれているツイート、ニュース記事の配信元が付与した見出しまたは記事全文画面へのリンクが含まれているツイートをそれぞれ関連ツイートとして取得する。ニュース記事に関連するツイートにおいて

表 1: データセット作成に用いたニュースリスト

トピック名	記事数	ニュースのサンプル
行政指導	1	ドワンゴへの行政指導について厚生労働省に質問を送ってみました
デング熱	32	東京都, デング熱など蚊媒介感染症対策を国へ緊急提案
		デング熱ってどんな病気? 解熱剤はアセトアミノフェンを
エボラ出血熱	36	発熱男性はエボラ陰性 都内30代, シエラレオネから帰国
		エボラに救世主 日本のアビガンが世界を救う
鳥インフルエンザ	8	日本にも上陸の可能性?
		鳥インフルエンザ A(H7N9) についての Q&A
ノロウイルス	24	ノロウイルスの有効な予防法は?
		93%の女性が「手洗い」と回答
子宮頸がん	10	子宮頸がんワクチン, 予防か安全性か 問題点をおさらい
危険ドラッグ	17	麻薬取締官 29人増員=危険ドラッグ対策で一政府
イクメンプロジェクト	3	イクメンになれないパパたちの悲痛な叫び: その3

は記事の内容が反復されている場合や、記事の見出しが含まれている場合が多いため、収集した関連ツイートからユーザが独自に書いた「コメント」部分を抽出する。このようなコメント部分を「コメントツイート」と呼ぶものとする。

B.2.2 データセット

本研究分担では、2種類のデータセットを用意した。1つ目は、2014年12月7日から2015年1月15日までの40日間に、Yahoo!ニュースのトピックスとして扱われた記事4078件を収集したデータセットである。「国内」「国際」「地域」「経済」「エンターテインメント」「スポーツ」「コンピュータ」「サイエンス」の8つのカテゴリから網羅的データを収集した。こちらを「網羅的データセット」と呼ぶ。

2つ目は厚生労働省に関連の深い話題についてニュース記事リストを生成し、そのリストに基づいて記事128件を収集したデータセットである。用いたトピックと

しては、「厚生労働省による行政指導」「デング熱」「エボラ出血熱」「鳥インフルエンザ」「ノロウイルス」「子宮頸がん予防ワクチン」「危険ドラッグ」「イクメンプロジェクト」である。各トピックの代表的な記事のタイトルと記事数は表1のとおりである。こちらを「厚労省データセット」と呼ぶ。

B.2.3 相関分析

感情的な言葉がある程、読者が共感しやすい[2]というこれまでの研究結果について、本実験のデータセットでも同様の結果が見られるか分析を行う。ここではデータ量が十分にある網羅的データセットのみを用いる。なお、本研究分担では、各文書の感情の度合いを感情度という数値で表現した。感情度とは文書中の感情表現の出現数におけるポジティブ、ネガティブの割合を表す数値であり、ある文書中の感情表現に占めるネガティブ感情の割合が高いほど0に近く、ポジティブな感情表現の割合

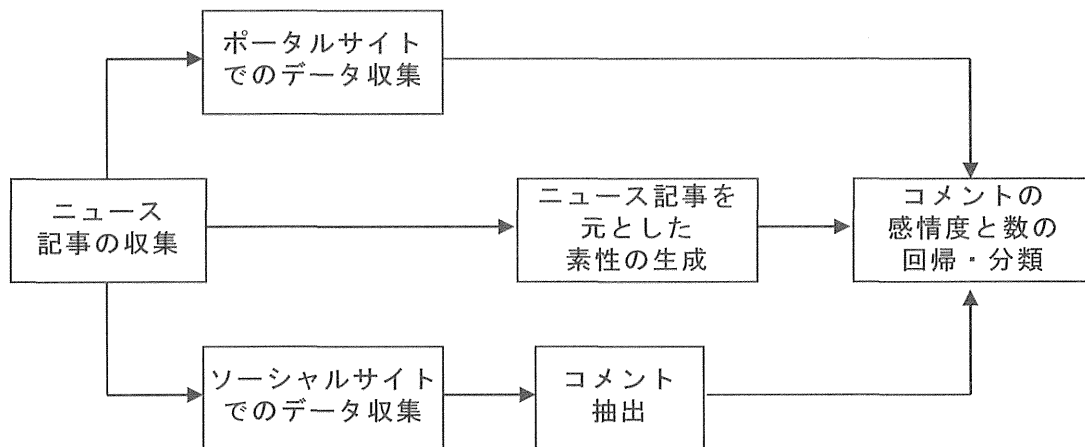


図 3: 提案手法のフレームワーク

が高いほど 1 に近い値となるように重み付けを行っている。

各ニュース記事及びそのコメント群について感情度を算出した後、それらについて相関分析を行う。

B.3 ニュース記事とその反響の大きさの相関

ニュース記事の内容がその反響に与える影響について分析を行う。従来のニュース記事の人気度予測研究からはニュースのカテゴリーや時間といった特徴のみならず、ニュースコンテンツに含まれる固有表現がニュースの人気度に影響を及ぼしていることがわかっている[1]。

既存研究を踏まえ、本研究分担でもニュース記事の言語特性に焦点を当てる。言語特性を対象とする理由は、既存研究により人気度に影響を及ぼすことがわかっているため、また、情報を発信する際に、得られた知見を活用する(たとえば、プレスリリースの書き方を変更する)ことが容易なためである。具体的には、固有表現や感情語以外にもニュース記事の文字数や、記号の有無、品詞の割合など言語的な要素を包括的に検討し、これらの言語特性がユーザの反応に与える影響を分析する。ユーザの反応分析にはポータルサイト“Yahoo!JAPAN”のニュースページとソ

ーシャルサイト“Twitter”の2種類のメディアから集めたコメントを使用する。なお、ニュースコンテンツのユーザの反応への影響度を捉える基準としては、ユーザのコメント数の伸びを用いる。手法のフレームワークは図3の通りである。

B.3.1 ニュース記事に対するコメントの収集

ニュース記事が受け手に与える感情の相関分析と同様の方法を用いる。

B.3.2 データセット

ニュース記事が受け手に与える感情の相関分析と同様のデータセットを用いる。

B.3.3 特徴量の生成

収集した記事から、特徴量を抽出する。本研究分担では、ニュースの表層的素性、言語的素性、時間的素性、環境的素性の大きく分けて4種類の素性を用いる。

表層的素性(SF) ニュース記事の持つ表層的素性として、Yahoo!ニュースのトピックスに掲載された際に付与されるカテゴリーと、ニュースの配信元の名称、そして独自の素性としてニュース記事

が掲載されているリンク先の種類 (Yahoo!ヘッドライン, Yahoo!ニュース BUSINESS, Yahoo!個人, Yahoo!政治, Yahoo!選挙, Yahoo!ニュース:新着雑誌記事, スポーツナビ, ネタリか)を用いる。

言語的素性 (LI) ニュース記事, ニュース見出し, ニュース関連リンクからそれぞれ素性を生成する。ニュース記事では記事に出現する単語 (bi-gram, tri-gram), 文字数, 品詞の占める割合, 固有表現, 客観性, 感情語, ポジティブ・ネガティブ・ニュートラルの各感情語の数, 感情度を素性として用いた。ニュース見出しからは, Yahoo!トピックス独自の見出しと, 記事の見出しの両方から, 文字数, 記号と数字の出現頻度, ひらがなとカタカナのそれぞれの割合, 固有表現, 感情語, ポジティブ・ネガティブ・ニュートラルの各感情語の数を, 最後に両見出しの類似度を計算し, 素性として加える。ニュースの関連リンクからは, リンク先につけられた名前とヤフー編集者が付与したリンク先の要約に用いられた言葉を基に, 感情語とポジティブ・ネガティブ・ニュートラルの各感情語の数の素性を生成する。

時間的素性 (TM) ニュース記事が Yahoo!のトピックスとして掲載された時刻, 曜日, トップニュースとして扱われた時間, トピックスとして扱われた時間を素性として用いる。曜日に関しては国民の祝日の他に年末年始(12/29・1/3)を祝日として扱うこととする。

環境的素性 (EV) 札幌・東京・名古屋・大阪・福岡の5都市における各日の昼の天気概況(06時~18時), 夜の天気概況(18時~翌日 06時), 最高気温, 最低気温の情報を気象庁発表の過去気象データから取得し, ニュースが配信され

た日付における最高気温と最低気温, 配信時刻の天候を素性とする。天気は「晴れ」「曇り」「雨」「雪・みぞれ・あられ」の4つを素性とし, 天気概況の文中に各素性名が含まれる場合, 素性値を1, 含まれない場合, 素性値を0とする。気温についてはデータから得られた日最高気温・日最低気温をそのまま素性値として使用する。また本研究分担独自の環境的素性として日経平均株価の記事配信日における終値, 終値 - 始値, 最高値 - 最安値の値を素性として採用する。

これらの特徴量の詳細を表5に示す。

B.3.4 機械学習による分類・回帰

収集した記事の特徴量とそのコメント数に機械学習を適用することで, コメント数の伸びに影響と与える記事の特徴量を明らかにする。問題を簡単にするために, 本研究分担では, ある記事に対してコメント数が伸びるか, 伸びないかという2値分類問題とみなす。コメント数に閾値を設定し, その閾値を超えるか否かを予測する問題である。

閾値としては, 予備実験の結果, コメントツイート数100という値を設定した。つまり, 各記事についてコメントツイート数が100以上になるか否かのラベルを付与した後, 記事のラベルと特徴量に対して機械学習を適用することで, その2値分類問題を解く分類器を生成する。特徴量の組み合わせを変更し, それらに対応する分析を生成し, それらの分類器の精度を比較することで, コメント数の伸びに影響をあたえる要素を明らかにする。

(倫理面への配慮)

本研究は, すべて公開情報を対象としたデータ解析を行う。研究成果の公開に際しては統計情報を中心とし, 意図しない個人

情報等が混入しうる形での情報公開を行わない。

C. 研究結果

B 節で提案した分析手法を用いて得られた分析結果について述べる。

C.1 ソーシャルメディア上での情報発信の現状

図 6 に政府・省庁アカウントのツイート数、図 7 にフォロワー数、図 8 に被 RT 回数、図 9 に 1 つのツイートが受ける被 RT 回数の期待値、図 10 にフォロワーが RT をする確率を示した。図 6 により、ツイッターを最も活用している機関は外務省(6,050 ツイート)であり、首相官邸(2,693 ツイート)、外務省柔らかツイート(2,242 ツイート)、経済産業省ホームページ更新情報(2,116 ツイート)、防衛省・海上自衛隊(2,050 ツイート)と続く。厚生労働省は全体の 12 位(734 ツイート)であった。

図 7 により、最もフォローされているアカウントは首相官邸(災害情報)であり、総務省/消防庁、防衛省などのアカウントが続く。厚生労働省は全体の 7 位(228,570 フォロワー)であり、ツイート数が少ない割にはフォロワー数が多いことが分かる。フォロワー数の多いアカウントは災害情報など、暮らしに密接したものが多いため、厚生労働省からの情報も食品の安全や健康情報など、暮らしに密接した話題が期待される。

図 8 から図 10 は、公式アカウントからの情報が「リツイート」という手段でどの程度拡散したのかを分析したものである。厚生労働省のツイートは 27,428 回(10 位)リツイートされ、1 つのツイートがリツイートされる回数の期待値は 37.3 回(13 位)である。これらの順位は、厚生労働省のツイート数やフォロワー数の順位とだいたい近い位置にある。しかしながら、厚生労働省

のツイートをを見た人が「リツイート」操作をする確率は 0.11%(33 位)であった。これは、注意喚起のために同じ情報を繰り返し発信するなど、厚生労働省が発信する情報の性質に依るところもあるが、情報の拡散力という観点では残念な結果と言える。厚生労働省よりもフォロワー数が多く、リツイート率の高いアカウント(例えば首相官邸や海上自衛隊)を参考にする必要であろう。

表 8 は、11 件のトピックに関するツイートの数を示したものである。例えば、「エボラ出血熱」を含むツイートは 2013 年 3 月 1 日から 2015 年 1 月 10 日までの期間で 2,715,763 件あったことが分かる。厚労省を含む数とは、これらのツイートの中で「厚生労働省」もしくは「厚労省」を含むツイートの数である。例えば、エボラ出血熱に関するツイートの中で 70,988 件(2.61%)は厚生労働省に関する言及があったことを示している。この数字は、各トピックにおいて厚生労働省が発信した情報や、厚生労働省の対応などが議論される頻度を示している。

この分析によると、「国民の皆様の声」「健康寿命をのばそう！アワード」「イクメンプロジェクト」などのトピックは厚生労働省の名前と一緒に語られることが多く、逆に「ノロウイルス」「健康づくり大キャンペーン」「鳥インフルエンザ」などのトピックは厚生労働省と一緒に言及されることが少なかったことが分かる。厚生労働省が発信していたツイートの代表例を以下に示す。

1. 4 月 19 日頃から国民生活基礎調査の準備調査が始まります。調査員が訪問した際には、ご回答をお願いします。詳しくはこちら→ <http://t.co/4AFFcczxcT>
2. 【中国へ渡航される方へ】中国で鳥インフルエンザ A (H7N9) が発生しています。海外では、不用意な動物との接触は避けてください。特に家さんの飼育場所

や、生きた鳥の販売所、食用に鳥を解体している場所などには立ち入らないようにしてください。

3. 【産科医療補償制度の申請期限が近づいています】産科医療補償制度は、重度脳性まひのお子様とご家族を支援する制度です。申請期限は満5歳のお誕生日までです。まずはご相談ください。詳しくはこちら→
<http://t.co/UlBv9kvchE>

4. 【イクメン企業アワード・募集中】イクメン企業アワードでは、こんな企業を募集しています。例えば、パパママランチ会の開催や専用 SNS を設けて、社内で情報交換。おかげで、子どものお弁当作りや学校行事への参加と、育児に積極的なパパが増殖中！
<http://t.co/uDk0E4wr7K>

5. マメ家族のインフル奮闘記★【1/第 15 話】大学受験も終わり、あとは結果を待つばかりのマメゾウです。今週は、「咳エチケット」についてお伝えします。＜咳エチケットとは＞1. 咳やくしゃみをする時はティッシュなどで口と鼻を覆う 2. 鼻をかんだティッシュはすぐにゴミ箱捨てる（つづく）

6. 「正しく知ろう、エボラ出血熱」スタート国民の皆様へ、エボラ出血熱について、知っておいていただきたいことを、Q&A 方式でお知らせします。つぶやきは、最新の情報に基づいて発信いたしますが、今後情勢が変化する場合もありますのでご了承ください。

これ以外にも様々なトピックに関するツイートがあったが、基本的には国民への依頼(ツイート 1)、注意喚起(ツイート 2)、キャンペーンや制度の広報(ツイート 3, 4)、インフルエンザやエボラ出血熱などの特定のトピックに関する解説シリーズ(ツイート 5, 6)に集約される。「マメ家族のインフル奮闘記」は Togetter でのまとめサイトも出来ており、読者からも好評を得ていたと推測される。次に、被リツイート数の

多いツイートを以下に示す。

1. 【9月1日から「労働条件相談ほっとライン」を開設】「有給休暇が取れない」「残業が多い」など労働条件でお悩みの方、お電話ください。平日夜間・土日に無料で受付。電話：0120-811-610 まで。携帯からも OK。詳しくは→
<http://t.co/HnoksZw5r2>
2. 【エボラ出血熱対策に関するメッセージ】エボラ出血熱への対応について、塩崎大臣から国民の皆様へのメッセージはこちら→ <http://t.co/8i75qIT98o>
3. 厚生労働省が販売したとしている「ワクチン債」の勧誘があるようです。厚生労働省では、「ワクチン債」を発行したという事実はありませんので、購入することのないよう御注意ください。
<http://t.co/C4F3hluGd9>

概して、厚労省発のキャンペーンや注意喚起に関するツイートがリツイートされやすい傾向にある。また、ツイート中に URL が含まれているなど、ウェブサイト上の情報へ誘導している点も特徴的である。このように、ウェブサイト上での広報に加えて、ソーシャル・メディアで情報の拡散を促す戦略は、ある程度の成功が確認できた。

表 8 に示したように、「国民の皆様の声」「健康寿命をのばそう！アワード」「イクメンプロジェクト」などのトピックは厚生労働省とよく一緒に言及される。これは一見すると良いことのように思われるが、実際には次のようなツイートが多い。

1. 厚生労働省「国民の皆様の声」募集送信フォーム <http://t.co/rq38vj8ZDh>
2. 「第 1 回 健康寿命をのばそう！アワード」 | 報道発表資料 | 厚生労働省”
<http://t.co/EdiMcpQEZL>

残念ながら、これらのトピックに関しては個人やマス・メディアで言及されること

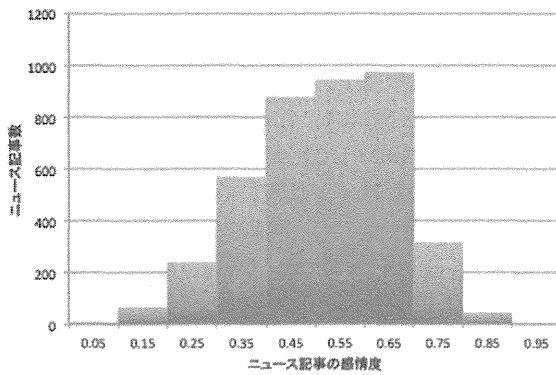


図 4: 感情度あたりのニュース記事数

が少なかったため、相対的にプロジェクトの運営元が言及されたと見るべきである。言い換えれば、プロジェクトの波及効果が少なく、独り立ちできなかつたと捉えることもできる。このように、厚生労働省が行った施策やプロジェクトの普及状況を、ソーシャル・メディア上での客観的な指標から推し量ることが可能である。

C.2 ニュース記事が受け手に与える感情の相関

各ニュース記事の感情度とコメントの感情度について、相関分析を行った。

まず感情度あたりのニュース記事数をまとめたヒストグラムを図 4 に示す。本実験ではニュース記事は感情度が 0.4 ~ 0.7 であるものが、全体の 69% を占め、特にニュートラルな感情(感情度:0.5)よりも少しポジティブな感情をもつ記事が Yahoo!ニュースのトピックスとして扱われる記事の中では多いことがわかる。網羅的データセット内におけるニュース記事の感情度とコメントツイートの感情度との相関を図 5 に示す。図 5 より、ニュース記事の感情にコメントを付ける読者も同じような感情を持ち易い傾向があることが読み取れる。なお、相関係数は 0.59 であった。

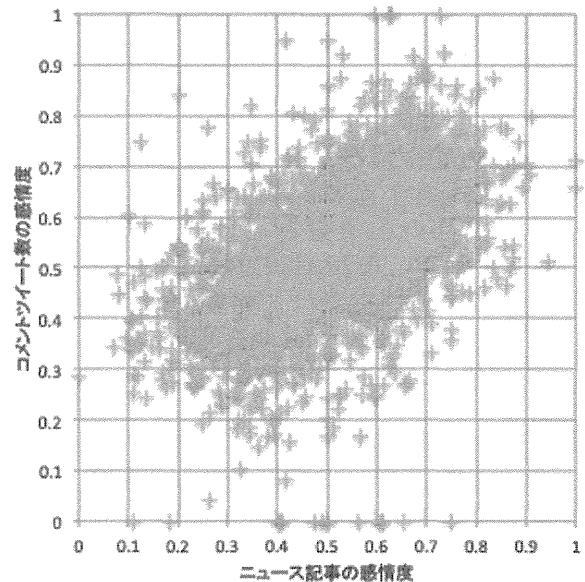


図 5: ニュース記事の感情度とコメントツイートの感情度の関係

C.3 ニュース記事とその反響の大きさの相関

各ニュース記事の特微量を用いてコメント数が多いと予測される記事、少ないと予測される記事の分析を行う。網羅的データセットを用いて、特微量の組み合わせを変えつつ、Support Vector Machine を用いて 5 分割交差検定により精度比較を行った。その結果を表 2 に示す。

ここで厚労省データセットを用いないのは、そもそもデータ数が少ないために交差検定を行うことが困難なためである。

表層的特徴(SF)・言語的特徴(LI)・時間的特徴(TM)・環境的特徴(EV) の内、言語的特徴(LI) が単独では最も影響力があるが、時間的特徴(TM) のみを用いた場合を除き、どの素性組み合わせにおいても F 値 0.65~0.70, 正答率も 60%後半~70%程度であった。

次に言語的な特徴を詳細に分類して精度を比較した結果を表 3 に示す。文字特徴(LI-Character)・固有表現特徴(LI-NamedEntity)・感情特徴

(LI-Sentiment) の内では文字特徴 (LI-Character) がコメントの伸びに最も寄与していることがわかり、全言語特徴 (LI)を用いた学習結果とほぼ同等の結果が得られた。

表 2: 関連コメントツイート数分類における特徴量の比較(全体)

特徴	F 値	正答率
SF	0.652	64.3%
LI	0.674	67.1%
TM	0.433	55.3%
SF+LI	0.675	68.3%
すべて	0.674	68.3%

表 3: 関連コメントツイート数分類における特徴量の比較(言語)

特徴	F 値	正答率
LI-Character	0.670	66.9%
LI-NamedEntity	0.663	63.9%
LI-Sentiment	0.636	62.3%

次に厚労省データセットにロジスティック回帰を適用し、分類に大きく影響する言語的特徴量を調べた。結果は表 4 のとおりである。表 4 より、コメント数の伸びに対して、ポジティブな影響を与える言語的特徴として、見出しに関する特徴量、特に見出しに含まれる評価語数が上位に位置づけられている。これより、見出しに評価語を多く含む方がコメント数が伸びやすい可能性がある。また見出しの文字数も長いほうが良いことがわかる。逆に、コメント数の伸びに強いネガティブな影響を与える言語的特徴として数値及び句点が上位に位置づけられている。これより、数値が多い記事や全体的に文が多い記事はコメントが伸びにくい可能性があることが読み取れる。

D. 考察

ソーシャルメディア上での情報発信の現

状分析においては、厚生労働省の Twitter アカウントとその他の政府・省庁系の Twitter アカウントの比較を行った。本分析では、被 RT 回数の平均値や期待値など

表 4: 分類に有用な言語的な特徴量の上位 5 個

ポジティブ	ネガティブ
negative (見出し)	単語：9
ネガ評価語数(見出し)	単語：.
neutral(見出し)	単語：5
ニュートラル評価語数(見出し)	単語：3
見出し文字数	単語：2

各アカウントのツイート全体に対する比較を行ったが、個別のツイートや時間帯など詳細な分析を行うことで、より有用な知見を得られる可能性は高い。定量的な分析だけでなく、情報の拡散力が高いアカウントについて人手で定性的な分析を行うことも有用であろう。また、政府・省庁系アカウントのみならず、自治体や企業のアカウントで情報発信力の高いアカウントについても分析も検討すべきであろう。

ニュース記事が受け手に与える感情の分析においては、各記事全体の感情度とコメントの感情度について相関分析を行った。結果として、記事と同程度のポジティブまたはネガティブな感情をコメントが示しやすい傾向があることを明らかにすることができた。しかしながらユーザの多くが記事に共感しやすいのか、記事に共感した場合にコメントをし易いのか、という点には議論が残っており、より詳細な分析が必要であろう。また、より詳細な分析として、記事の見出しや本文の前半、中盤、後半など、記事を細かい要素に分解した際に、どの要素の感情度とコメントの感情度が相関するかを調べることも考えられる。

ニュース記事とその反響の大きさの分析

では、ニュース記事の言語的特徴がユーザのコメントに少なからず影響を与えていることを示唆する実験結果を得た。ただし、ニュースのカテゴリによって、どのような特徴がコメントに影響を与えるかは違いが生じると想定される。厚生労働省に関わるトピックにおいて反響により大きな影響を与える要素を明らかにするためには、厚生労働省のプレスリリースや厚生労働省に関連する記事及びそれに対する反応を継続的に収集していく必要があるであろう。

E. 結論

本研究分担は、ウェブやソーシャルメディアでのリスクコミュニケーションにおいて、より効果的な情報発信手法を開発することを目的として、ソーシャルメディアアカウントの投稿やウェブ上のニュース記事に対するソーシャルメディア上のユーザの反応を分析した。具体的には、ソーシャルメディア上での情報発信の現状分析、ニュース記事が受け手に与える感情の相関分析、ニュース記事とその反響の大きさの相関分析を行った。分析を通じて、ソーシャルメディア上の情報発信について定量的な効果測定が可能であること、ニュース記事の感情が受け手の感情に影響をあたえること、ニュース記事に含まれる言語的特徴が反響の大きさに影響しうることを明らかにした。

ソーシャルメディアが普及した現在では、一般人までもが独自に情報を発信する状況であるため、今までのようなリスクコミュニケーションのアプローチが有効でないこ

とが想定される。しかし、同時に、情報の受け手の反応を観測することができるため、それをフィードバックとして、リスクコミュニケーションにおける情報伝達の手法をよりよいものに修正していける可能性がある。今後は、本研究分担で得られた知見を活かして、ウェブやソーシャルメディアを用いたリスクコミュニケーションにおける情報発信の手法を現状に即したものに継続的に改良していくことが期待される。

F. 研究発表

1. 論文発表

なし

2. 学会発表

なし

参考文献

- [1] Roja Bandari, Sitaram Asur, and Bernardo A Huberman. The Pulse of News in Social Media: Forecasting Popularity. In ICWSM, 2012.
- [2] 内藤和宏, 榎堀優, 梶田将司, 間瀬健. Twitter コメントに含まれる感情語がイベント印象に与える影響の評価. インタラクション 2012, pp. 871-876, 2012.

表 5: 機械学習に用いた特徴量

素性	定義	値
表層的素性 (SF)		
カテゴリー	{ 国内, 国際, 地域, 経済, エンターテインメント, スポーツ, コンピューター, サイエンス }	0-1
記事の配信元	ニュース記事の配信元の名称	0-1
記事のリンク先	{ Yahoo! (ヘッドライン, ニュース BUSINESS, 個人, 政治, 選挙, ニュース:新着雑誌記事), スポーツナビ, ネタリか }	0-1
言語的素性 (LI)		
記事 (LI-AR)		
N-gram(Character)	$DF \geq 2$ のバイグラム, トリグラム	0-1
TF-IDF(Character)	$DF \geq 2$ の TF-IDF 値	Float
文字数 (Character)	記事の文字数	Int
品詞 (Character)	記事中の各品詞の割合	Float
固有表現 (NamedEntity)	記事に含まれる固有表現 (人名・組織名)	0-1
客観性 (Sentiment)	記事が主観的であるか客観的であるかを表した 0 から 1 までの値	Float
感情語 (Sentiment)	記事に含まれる感情語	0-1
感情語の数 (Sentiment)	記事に含まれるポジティブ・ネガティブ・ニュートラルの各感情語の数	Int
感情度 (Sentiment)	記事がどれだけ感情的であるかを表した 0 から 1 までの値	Float
見出し (LI-HE) (Yahoo!トピックスの見出し (LI-HE-y)・記事の見出し (LI-HE-n))		
文字数 (Character)	見出しの文字数	Int
記号・数字 (Character)	記号 {!, ?, 「, 「, &, =, ,, , <, 「, " , % }, 数字の出現回数	Int
ひらがな・カタカナ (Character)	見出しに含まれるひらがな, カタカナそれぞれの割合	Float
固有表現 (NamedEntity)	見出しに含まれる固有表現 (人名・組織名)	0-1
感情語 (Sentiment)	記事に含まれる感情語	0-1
感情語の数 (Sentiment)	記事に含まれるポジティブ・ネガティブ・ニュートラルの各感情語の数	Int
見出しの類似度 (-)	Yahoo!トピックスの見出しと記事の見出しの Jaccard 類似度	Float
関連リンク (LI-RL)		
感情語 (Sentiment)	記事に含まれる感情語	0-1
感情語の数 (Sentiment)	記事に含まれるポジティブ・ネガティブ・ニュートラルの各感情語の数	Int
時間的素性 (TM)		
曜日	Yahoo!ニュースで配信された曜日 { 月, 火, 水, 木, 金, 土, 日, 祝 }	0-1
時刻	Yahoo!ニュースで配信された時刻 { 0 - 23 }	0-1
配信時間	Yahoo!ニュースのトピックスとして, トップニュースとして扱われた時間	Int
環境的素性 (EV)		
天気	配信日昼/夜の札幌・東京・名古屋・大阪・福岡の天気 { 晴れ, 曇り, 雨, 雪・みぞれ・あられ }	0-1
気温	配信日の札幌・東京・名古屋・大阪・福岡の最高気温・最低気温	Int
日経平均株価	配信日の終値, 終値-始値, 最高値-最安値	Float